

Guidelines for the Use of Social Media Data in Research

The University has produced these guidelines as part of a commitment to the highest standards of research, underpinned by the quality of the research process, providing a framework through which those involved in the research process can address the ethical implications of utilising social media data in research. Social media as a research tool and data source presents new opportunities for researchers, but also new challenges which require careful consideration throughout the lifetime of the project. The University recognises the difficulty with producing a set of definitive rules, which may become out of date as technology changes. As such, these guidelines provide advice on factors for consideration when using social media content in research, as well as setting out the position of the University on key issues.

When contemplating a research project making use of social media, the main guiding principle to consider is the same as with offline research – the avoidance of harm. In terms of social media data specifically, the University understands social media users as **human participants**. The University places trust in the ability of researchers to assess their own work, with assistance from colleagues and subject-level ethics committees where relevant, to identify where ethical review is necessary, and provide guidance where needed. It is strongly advised that researchers spend time getting to know the social media platform they intend to use, both in terms of policies and culture. Importantly, however, the culture of platforms must not override the University's ethical framework.

Research making use of social media is a wide-ranging field, and so will be guided by a variety of considerations. This guidance explores five of the most prominent issues noted in social media research. These are: legal considerations, avoidance of harm, data integrity and management, privacy, and consent. The key elements of each area are summarised below.

Consideration 1: Legal considerations

Legal considerations refer to (a) the Terms and Conditions of a platform, and (b) legislation surrounding GDPR and data protection. All research at York must be conducted in compliance with legal and regulatory frameworks; this is, however, a complex and potentially difficult area to map out with regard to social media data. Researchers must have a working knowledge of the Terms and Conditions of any social media platforms or apps they use, and be mindful of potential legal issues.

Potential issues can arise when negotiating with social media companies and platforms. Such groups may wish to limit or control access to data and subsequent narratives, and researchers must consider throughout the impact this has on the research they produce and their academic freedom to do so. If you intend to negotiate with platforms or companies in order to access data, please contact the [Research & Knowledge Exchange Contracts team](#).

Information generated by, on, and for social media networks may be [personal data](#) if the user can be reasonably identified, and the University [Data Protection team](#) is available to assist with clarifications. Researchers must familiarise themselves with the requirements of GDPR, and ensure compliance.

Consideration 2: Duty of Care

In line with our commitment to the avoidance of harm, social media researchers should consider their duty of care to those involved in the research process. With social media research, this most often refers to potential harm towards participants, researchers, and the wider community. Particular consideration must be paid to the welfare and interests of vulnerable individuals, although it should be recognised that identifying vulnerability is a complex process on social media. It is important to note that studies taking place through social media may find it difficult to effectively debrief participants or ensure appropriate care following research. Researchers should consider how best to mitigate this risk, and ensure that social media research is in the best interests of both the research project and its participants. When dealing with such data, it is good practice to seek consent where quoting directly, or to paraphrase, and, where there are high levels of potential risk, consider whether social media is the best method of data collection.

Specific consideration must be given to the potential for data to be supplied by underage individuals. Researchers should consider ways in which this can be addressed – researchers are not expected to completely eliminate every risk, but must mitigate as far as possible, and be aware of shortcomings or potential issues.

Consideration 3: Data Integrity and Management

Social media data can be difficult to verify – users may lie about their age, location, job, or any number of other characteristics. Researchers must be aware of this issue and address this difficulty where relevant. It is not advisable to understand users as the ‘general public’, due to inequalities in access to the internet, and researchers should consider how to foster diversity (where relevant) in their sample.

Researchers must refer to the University policy on [Research Data Management](#) to ensure such data is handled in line with regulations. Completion of a Data Management Plan will help researchers think through this process, and identify potential concerns. The completion of a [Data Protection Impact Assessment](#) is recommended and in many cases will be necessary, particularly when planning to monitor a ‘public’ space (see next section) on a large scale.

Consideration 4: Privacy

In order to conduct research ethically, it is important to consider the extent to which users *believe* a website or app is private, alongside the ‘reality’ of the site as a public place. Certain forums or groups, for example, may technically be free to access and openly available, but the members and contributors could still feel violated if their data and comments were used without consent. It is important to be aware that material from social media platforms may

be easily searchable, therefore making users identifiable and potentially putting them in a vulnerable position. The data collected and stored should be proportionate to the needs of the study, and specific guidance regarding how and when to anonymise data must be sought from the relevant ethics committee in the first instance – please be aware there may also be journal-specific policies that must be adhered to.

Consideration 5: Consent

When considering consent from an ethical standpoint, researchers should take into account, amongst other factors:

- the views of gatekeepers and users;
- the level of engagement with users;
- whether the observation of a non-public space is the most appropriate approach for research; *and*
- how informed consent can be gained practically.

Importantly, the simple act of agreeing to the Terms and Conditions of a website does not constitute *informed* consent, as it cannot be ensured participants have fully read or understood the agreement.

Low-risk research involving no direct contact with participants or personal details may not require informed consent. Research using social media data has the potential to fall into these categories; such situations must, however, be fully justified, and are not exempt from the requirement of ethical review to ensure this is the case.

Researchers shall take time to consider how withdrawal works in the context of social media research – does deleting a post, for example, count as withdrawing data? Could a researcher realistically become aware of this, and how? Similarly, there is a risk with social media research that individuals may be difficult to contact, or may not respond when contact has been made. Researchers must have clearly defined processes on how they seek consent and become aware of, and adjust to, possible withdrawal of data from the project.

These guidelines do not aim to restrict topics or areas of study, and the University recognises the potential value of social media data and content; such resources must, however, be used effectively and ethically. Subject level Ethics Committees should remain the first point of contact when seeking ethical approval, and queries shall be addressed to these in the first instance.

These guidelines apply to all those undertaking research under the University's auspices as defined in the [Code of Practice on Research Integrity](#). This guidance sits alongside other policies within the University, including the [Policy on Research Data Management](#), the [Policy and Procedure for the Investigation of Allegations of Research Misconduct](#)¹. This guidance is

¹ This list is not exclusive, and researchers should ensure they have familiarised themselves with the full range of University policy prior to beginning research.

also informed by existing guidance within the sector². Researchers should ensure familiarity with applicable University policy, including that governing the ethical review process – importantly, ethical approval must be in place **prior to beginning any research**.

Please note that examples given in the following guidance are simply for the purpose of illustrating principles. All research conducted under the auspices of the University of York must undergo appropriate ethical review and be assessed individually as set out in the [Code of Practice and Principles for Good Ethical Governance](#).

² See Appendix 2 'Further Reading'.

Consideration 1: Legal Considerations

A researcher is planning an examination of tweets related to political protest movements within the UK. She intends to collect aggregate-level data showing the usage of particular hashtags related to the movement in question, and hopes to quote some specific tweets to illustrate key points. She is aware that there will likely be some risks associated with the project, as well as potential legal ramifications, and wants to be sure she mitigates these as far as possible.

All research at York must be conducted in compliance with legal and regulatory frameworks; this is, however, a complex and potentially difficult area to map out with regard to social media data. The University [Research & Knowledge Exchange Contracts team](#), as well as the University [Data Protection Officer](#), are available for guidance in these areas.

1.1 Relevant Terms and Conditions

Our researcher should begin by ensuring she has a working knowledge of the Terms and Conditions of the platforms she intends to use – in this case, Twitter. Twitter's Terms of Service prohibit both the use of data-scraping tools on the platform, and the modifying of tweets when used in secondary material. These Terms and Conditions might conflict with those of journals, funders, or the University itself, and will need to be balanced appropriately.

Our researcher might decide to negotiate access to the data [directly with Twitter](#) in order to avoid potential legal complications. Potential issues can arise when negotiating with social media companies and platforms. Such groups may wish to limit or control access to data and subsequent narratives, and researchers should consider throughout the impact this has on the research they produce and their academic freedom to do so. Obtaining data from such platforms may take time or require extra steps in the research process which must be planned for. The University **explicitly forbids the granting of retrospective ethical approval**, and as such it is imperative this is negotiated in advance and signed off by an ethics committee. Further, whilst such companies may be the legal owners of the data, from an ethical standpoint explicit consent from users themselves as authors of specific posts may be expected by the University (see Appendix 1). Researchers intending to negotiate with platforms or companies in order to access data are advised to contact the [Research & Knowledge Exchange Contracts team](#).

As noted above, our researcher must also consider the fact that Twitter prohibits the modifying of tweets when used in secondary material, and requires that quoted tweets are credited to their author. This is a particularly thorny issue in terms of research ethics, as the anonymization of participants would most likely be the default position required by an ethics committee. How these legal and ethical requirements can best be balanced will need to be considered on a case by case basis to ensure steps taken are proportional and suited to the needs of the research. It might be that our researcher decides to paraphrase tweets rather

than quoting them directly, to ensure both compliance with Twitter's terms and conditions and the full anonymization of participants.

1.2 Data Protection

Our researcher will also need to consider the ways in which [personal data](#) can be protected in the research project in line with [GDPR requirements](#) as well as local guidelines where applicable³. [University resources](#) are available to assist staff with this, as is guidance from the [Information Commissioner's Office](#).

As she is using aggregate data, there is recognition within the GDPR legislation that it is not always practicable to contact each social media user to seek consent, and in cases where this would “involve disproportionate efforts”, consent may not be necessary⁴. However, our researcher will need to fully justify this choice as part of the ethical approval process. To ensure her handling of data is robust, a [privacy notice](#), outlining the project, uses of data, and the rights of participants shall be prepared and displayed in a noticeable location. The specifics of this will depend on the type of social media platform in question.

A template Privacy Notice for Social Media Research is appended to this document (see Appendix 1), which provides a generalised outline of how and why social media data is used for the purposes of research. This template privacy notice is not an acceptable substitute for a project-specific privacy notice; **an individual notice should be prepared for each project**. For assistance in drafting a privacy notice, researchers are advised to contact the University [Data Protection Team](#).

Our researcher should also consider the use of a [Data Protection Impact Assessment](#) (DPIA), in order to analyse the data processing activity itself and minimise any risk associated. This requires researchers to evidence the necessity of their work, identify risks, and offer mitigating measures. Completed DPIAs must be signed off by the University Data Protection Officer. Further information on DPIAs is provided by the Information Commissioner's Office, available [here](#).

³ See [here](#) for guidance on conducting research overseas. The University [Research & Knowledge Exchange Contracts team](#) and the [Data Protection](#) team are also available to provide guidance in this area.

⁴ GDPR [Article 14.5\(b\)](#)

Consideration 2: Duty of Care

A researcher wants to conduct a research project looking at experiences of online support groups for individuals dealing with eating disorders (ED). The researcher has identified some key online forums and intends to post his questionnaire on these for users to complete if they wish. He is mindful of his responsibility to avoid causing harm wherever possible, and knows this is a particularly sensitive area. This requirement must be considered, and risks mitigated, in advance of data collection. Whilst the majority of research will carry some potential risks, the ways in which this might manifest on social media can be unique and should be fully considered. Social media users might feel exploited or resentful of the extra attention that research involves, for example, or they might be put at risk by the inclusion of certain posts or comments in research. There is precedent for individuals losing their jobs or social support networks as a result of online activities, and research engaging in such areas should be mindful of the potential effects of reproducing and preserving postings in their work. Our researcher first decides it is key to spend time familiarising himself with the culture of the forums in question, to enable an understanding of the dynamics at play on the website.

2.1 Appropriate Handling of Participants and Incidental Findings

Studies taking place through social media may find it difficult to effectively debrief participants or ensure appropriate care following research; this is particularly pertinent in the case of research on sensitive topics. In some cases, researchers may decide social media is not the appropriate avenue for the research. Prior to beginning research, some consideration should be given as to how incidental findings might be handled, ensuring reporting routes are identified and communicated both to participants and the research team. For example, specifying that references to abuse in social media posts are reported in line with safeguarding procedures. It is good practice that, where research takes the form of a question set on a sensitive topic, relevant resources are signposted at the beginning and end of the survey.

Our researcher should also consider the possibility that underage or vulnerable individuals might inadvertently be included in his dataset. For example, even where social media platforms have a minimum age for membership, young people can and will sign up anyway, meaning usage of an age-limited platform is not necessarily a proxy for confirming the age of participants. As he intends to conduct research via a questionnaire which asks for informed consent, the potential involvement of minors is especially problematic. Our researcher recognises this limitation but is unsure how to combat the possibility. He decides to address this directly in his ethics application, and speaks to the Chair of his subject-level ethics committee in advance of submitting his proposal to discuss possible ways in which to mitigate the risk. Researchers are not expected to completely eliminate every risk, but must mitigate as far as possible, and be aware of shortcomings or potential issues. Taking steps to, for example, flag certain keywords in responses which might indicate a user is underage or in

other ways vulnerable (such as references to drugs, alcohol or other situations which might diminish an individual's ability to fully consent) would demonstrate a real attempt to avoid using data from participants who cannot fully consent.

For the researcher in question, social media is inherent to his research, and so his focus will be on accessing information as ethically as possible. In some cases, however, researchers might like to consider whether social media is the best avenue for data collection. Researchers shall refer to their subject-specific guidelines where possible and submit proposals to Departmental Ethics Committees for individual consideration.

Consideration 3: Data Integrity and Management

A researcher wants to examine attitudes to proposed changes to a law, and decides to use social media data to do so. She intends to recruit via a range of platforms including Reddit and Facebook, and will analyse the data by key indicators such as location, age, and sex. She would like to quote relevant posts from the websites as well, but is unsure how to do this ethically and with respect to data management requirements. She will need to think through how her data is collected and managed in line with University requirements and research integrity.

3.1 Verifying Data and Handling Limitations

She is aware that such data can be difficult to verify – social media users may lie about their age, location, job, or any number of other characteristics. In some cases this may have implications for the avoidance of harm principle, which is of central importance to University ethical oversight⁵. For example, she must consider in advance if there is a reasonable and proportionate way to mitigate this risk, and how she might become aware of underage users in particular.

Samples consisting of data from social media platforms can be hard to control. A study looking at posts regarding a certain topic may struggle to limit potential participants by geography or age. This may be an issue when recruiting participants from a variety of platforms, resulting in a mixed-sample which will require additional work to ensure access to information is consistent and that the sample is internally coordinated.

The use of direct quotes is a debated area in research ethics, and researchers will need to consider the full scope and range of their research, including where and how they intend to publish, to ascertain the appropriateness of direct quotes. For more on this, see C4: Privacy.

⁵ See C2 for more of the Avoidance of Harm and Duty of Care

Our researcher is aware of inequalities in access to social media and the internet, and the average user of these resources may differ from the average person in society⁶. She decides that this should be clarified where possible to make clear that such participants cannot be understood as the 'general public'. When thinking through how best to identify participants, she decides that recruiting via her own personal accounts has too strong a potential to result in a sample biased towards her own social circle. She decides to create a new account specifically for the project, however she is aware that this will require a time investment to build up contacts. This is discussed further in C4.

Care must be taken not to misrepresent the meaning or context of a post, whether unintentionally, recklessly or maliciously, and researchers should consider ways in which to mitigate such risks. Simple engagement with a 'Trending Topic' on Twitter, for example, should not be assumed to be a positive response – it may be that certain users are using the Trend to raise other issues, critique the topic, or simply be sarcastic. The online environment can be nuanced and complex, and it is important that researchers are aware of such possibilities. It is important that researchers have the capacity to be self-critical and willing to fully consider whether they are informed enough to be using social media data and platforms as data sources. This is particularly the case with international platforms, where supervisors or other individuals involved in a project may not fully understand the site in a legal or ethical sense.

3.2 Managing Data

Our researcher is aware of the need to manage her research data appropriately and in line with the University policy on [Research Data Management](#). This policy remains relevant at all stages of the project when collecting, using, and eventually archiving or disposing of the data. She decides to complete a Data Management Plan to assist her in thinking through this process, and identifying potential concerns. She also decides to make use of University [resources](#), and signs up to attend a training session looking at [Research Integrity and Ethics](#), as well as a session on [Research Data Management](#). In terms of open access, she refers to the Data Management Policy again and decides to speak to the Data Protection Officer as advised to understand how open access requirements affect her work. If she wishes, further advice could also be sought from the Library [Research Support team](#).

Another area our researcher will need to consider is the method through which she collects data, in particular the use of web-services to 'scrape' information from platforms. She is responsible for understanding how the data is processed and stored, and must be sure to

⁶ See Taylor, J., and Pagliari, C. (2018). "Mining Social Media Data: How are Research Sponsors and Researchers Addressing the Ethical Challenges?". *Research Ethics* 14(2): 3 for a brief outline, and Munson et al. (2013). "Sociotechnical Challenges and Progress in Using Social Media for Health". *Journal of Medical Internet Research* 15(10): e226 for a discussion of this in relation to health data.

include a consideration of web-scraping services in her ethical review proposal⁷. This will include the ability of such services to capture metadata (such as geospatial data or data relating to special categories), which should be considered specifically as part of this review. Further, the decision to recruit participants through social media requires careful thought as to the potential dangers associated with using a third party to collect data, particularly the possibility that such services do not appropriately secure or manage the data. Data may also be stored outside the EU, which should be addressed in a [Data Management Plan](#). The University offers approved tools such as [Qualtrics](#) to aid in the collection of data, which can help ensure data is handled appropriately.

Consideration 4: Privacy on social media

A researcher is planning a piece of work examining responses to a recent TV show, analysing levels of engagement on Twitter (specifically, the numbers of tweets using specific hashtags or words). He believes this is a relatively non-offensive topic, as he is using aggregate data from a 'public' platform, and so does not think privacy needs to be considered in depth. The idea of 'public' in terms of social media platforms is, however, one in need of greater interrogation; in order to conduct research ethically, it is important to consider the extent to which users *believe* a website or app is private, alongside the 'reality' of the site as a public place. This is distinct from legal concerns surrounding platform terms and conditions; certain websites and platforms, for example, may technically be free to access and openly available, but the members and contributors could still feel violated if their data were used without consent.

4.1 Assessing Context and Expectations

Our researcher must consider also the context surrounding their data; a tweet making use of a hashtag to engage in a wider topic may have a different goal or intended audience than a similar tweet without any hashtag. In a broader sense, nuances on platforms themselves can alter our perception of the private and public - platforms such as Facebook and Twitter restrict the content that can be seen when a user is 'logged out', even on 'public' profiles. How might this alter your understanding of 'public' in such a space? Researchers might find it useful to think through how similar dynamics in *offline* semi-restricted areas might be handled - ticketed events, for example, require some permission to access in the same way as an online forum. Where online groups or forums have a moderator or similar gatekeeper, it is good practice to contact them prior to commencing research.

⁷ For an overview of tools to retrieve and analyse social media data, see: Ahmed, W. (2019). "Using Twitter as a data source: an overview of social media tools". *London School of Economics: LSE Impact Blog*, available [here](#).

It is important to be aware that material from social media platforms may be easily searchable, therefore making users identifiable and potentially putting them in a vulnerable position - this risk is most prevalent when directly quoting users, however it can also apply to aggregate data. When quoting directly, it is likely that informed consent will be needed unless otherwise specified. A proportional approach to risk is advisable, and this is best done on a case-by-case basis, considering individually the intended participants and methods of the research. In the example given here, for instance, we might assume responses to a TV show is relatively low-risk, and without further information this would be a reasonable starting point - imagine, however, that the research will involve discussion of topics such as terrorism and crime. Does this change our initial assumption?

The particular format of data complicates this point further – text-based information differs from images, audio, or visual, and researchers must consider thoroughly prior to commencing research how they will collect and store data in a way that allows for anonymisation when necessary. Whilst anonymisation may be technically achievable for a single video, image or post, this does not mean the original content is not still traceable. Legal consideration is also needed here, as anonymising or blurring faces will require manipulating the original file. Large datasets also represent a challenge in this area, and researchers may find it difficult to anonymise individual extracts when these are included in publications. Researchers should seek advice and guidance from their subject-level ethics committee, as well as the [Research & Knowledge Exchange Contracts team](#) and the [Data Protection](#) team. Researchers are encouraged to speak to their Departmental Ethics Committee in the first instance to think through such issues.

It is recommended that researchers consider what it may be reasonable for users to expect – will they, for example, expect their posts to be used for research and then stored long-term in a public archive? Researchers should consider the archiving and retention of data more broadly than just GDPR requirements. Recent discussions surrounding open data might result in social media data being publicly available for longer than users may assume. The data collected and stored must be proportionate to the needs of the study, and specific guidance regarding how and when to anonymise data shall be sought from the relevant ethics committee in the first instance. This is with recognition that there will be discipline-specific issues and considerations, and it is beyond the scope of this document to outline these completely. Researchers should be aware that there may be journal-specific policies that must be adhered to regarding the reproduction of quotes and use of data⁸.

A final dimension of privacy that should be considered is that of the researcher and research team. Certain recruitment methods may mean researchers are searchable by participants,

⁸ For example, Taylor & Francis requires “the twitter author must be contacted and their permission sought” in order to use their content. See [here](#) for further information

particularly where privacy settings are not sufficiently secure. Some researchers choose to set up separate accounts specifically for the purposes of research, however this will require some practical consideration of how contacts might be collected and maintained.

Consideration 5: Consent

A researcher is planning a project using data from Facebook examining involvement in local history groups, and wants to know how to approach ‘consent’ when doing this work. She knows that the ethics of seeking consent go beyond the legal considerations discussed in Consideration A. Importantly, the simple act of agreeing to the Terms and Conditions of a website does not constitute *informed* consent, as it cannot be ensured participants have fully read or understood what they have agreed to⁹.

There are a number of things our researcher should consider when thinking about consent, including but not limited to the views of gatekeepers and users; the level of engagement with users; whether the observation of a non-public space is the most appropriate approach for research; and how informed consent can be gained practically. The specifics of these areas will differ between projects. In regards to our researchers proposed study, she plans to contact the moderator(s) of the Facebook groups and pages in question to start an initial conversation – where possible, this is good practice.

5.1 Assessing the Requirements surrounding Consent

Users’ perceptions of privacy can differ from the ‘reality’ of social media as a public space. Ethically speaking, it is important to consider how users understand or interpret an online space; this can vary between an expectation of public attention on, for example, a public Twitter account, and the perception that interactions on Facebook will be mostly limited to ‘real world’ contacts. When using appropriately anonymised data drawn from ‘public’ platforms, our intuition surrounding consent may differ when dealing with platforms felt to be ‘private’¹⁰. The British Psychological Society Guidelines for Internet-Mediated Research (2013) state:

“Where it is reasonable to argue that there is no likely perception and/or expectation of privacy (or where scientific/social value and/or research validity considerations are deemed to justify undisclosed observation), use of research data without gaining valid consent may be justifiable.”

⁹ Information Commissioner’s Office (2017). *Big data, artificial intelligence, machine learning and data protection*, Version 2.2. (ICO: England): 32.

¹⁰ ‘Public’ to be considered with reference to section 4a

Low-risk research involving no direct contact with participants or personal details may not require informed consent. Research using social media data has the potential to fall into these categories; such situations must, however, be fully justified, and are not exempt from the requirement of ethical review to ensure this is the case. Our researcher would find it useful to consider what 'low-risk' means in the context of her work – are comments relating to, for example, vintage photographs of streets in the local area of a different risk-level than comments relating to previous conflict in the community or incidents of illegal activity. It may be the case that informed consent is required for some aspects of participation but not others – quoting directly may require the full agreement of the author, whilst paraphrasing or using aggregate data (e.g. the number of comments) may not.

Our researcher decides to spend some time thinking about how social media users might respond to the research with regards to consent. She knows there is some evidence to suggest, for example, that members of the public are more comfortable with their social media data being used for academic research than for political or targeted market research¹¹. In these cases, there may be a perception that the surrendering of personal data is worth it for the 'trade-off' of research which offers a public good. This assessment is, however, inherently subjective, and it can be difficult to judge how large groups of users will feel; again, such a judgement must be made on a case by case basis. She decides to approach the question from a different angle, and consider how she would feel if her data were used for research; where would she draw the line? What would she be comfortable with? And what would she reasonably expect to be notified of?

5.2 Developing an Appropriate Process

Our researcher decides to seek consent for the use of direct quotes from users. To do so effectively, she will also need to consider how withdrawal works in the context of social media research – does deleting a post or comment, for example, count as withdrawing data? What might happen if a contributor is no longer on the platform? Could a researcher realistically become aware of this, and how? She decides to draw up a draft procedure, in conversation with relevant ethics committee members and seeking guidance from the Research Strategy & Policy Office where needed, as to how such contact can be managed, including where a participant does not respond.

Researchers should have clearly defined processes surrounding consent and possible withdrawal of data from the project. Part of the ethical approval process for our researcher will be identifying practical measures through which consent can be sought, and justifying the specifics of these. It is recognised that in the case of aggregated data this may be difficult -

¹¹ See Golder, S., et al (2017), "Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review" in *Journal of Medical Internet Research* for a discussion of this topic, available here: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5478799/>

researchers should consider available resources, and be careful not to make assurances that cannot be met in reality. For example, in the case of aggregated anonymised data, a researcher may struggle to remove a specific individual after a certain date in the study, and this should be made clear when appropriate. Similarly, where a platform specifically allows anonymous posting (as opposed to pseudonymous), this can make seeking consent impossible. Such eventualities must be considered when preparing an ethics proposal to ensure these are accounted for.

In cases where our researcher wants to use data without asking for consent prior to processing, she prepares a publicly available privacy notice outlining how data will be used, for what purpose, and how it is being or will be stored (see Appendix 1 for more).

Appendix 1: Template Privacy Notice

General Privacy Notice: Social Media Data in Research

This is a general privacy notice for people whose data is collected during research projects focussed on users of social media and the content of information posted on social media applications.

It sets out the general ways in which the University of York gathers, uses, stores and shares your data. It also sets out how long we keep your data and what rights you have in relation to your data under the General Data Protection Regulation (GDPR).

For the purposes of this privacy notice, University of York is the [Data Controller](#) as defined in the General Data Protection Regulation. We are registered with the Information Commissioner's Office and our entry can be found [here](#). Our registration number is: Z4855807.

Where do we get your data from?

Unless otherwise specified in relation to a particular research project, data are collected directly from publicly accessible social media sources, including but not limited to Twitter, Facebook, Reddit, etc.

What data do we have?

Unless otherwise specified in relation to a specific research project, aggregate-level data and specific social media posts.

How do we use your data and what is our legal basis for processing your data?

Data will be processed to advance ethically-approved research projects at the University and benefit society as a whole. In line with our charter which states that we advance learning and knowledge by teaching and research, the University processes personal data for research purposes under Article 6 (1) (e) of the GDPR: *"Processing is necessary for the performance of a task carried out in the public interest."*

Where special category is also processed the University will typically also rely on Article 9 (2) (j) of the GDPR: *"Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes in accordance with Article 89 (1)."*

Who do we share your data with?

Depending on the nature of the research project, data may be shared with third parties such as:

- Researchers and collaborators at other institutions;

- Grant funding agencies;
- Data warehouse and archiving facilities

Additional information about data sharing will be identified in privacy notices for specific research projects.

Further information

For further information on how the University uses your data and your rights under data protection legislation see, <https://www.york.ac.uk/records-management/dp/your-info/generalprivacynotice/>.

Appendix 2: Further Reading

Items listed on the reading list are presented to enable further thought and consideration of key issues. These items may not align exactly with University policy and in some cases may predate GDPR - it remains the responsibility of researchers to ensure correct procedure is followed.

Ahmed, W. (2017). "[Using Twitter as a data source: an overview of social media tools \(2017\)](#)". *London School of Economics: LSE Impact Blog*.

Ahmed, W. (2019). "[Using Twitter as a data source: an overview of social media tools \(2019\)](#)". *London School of Economics: LSE Impact Blog*.

Ahmed, W., Bath, P. and Demartini, G. (2017). "[Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges](#)". In Woodfield, K., (ed.). *The Ethics of Online Research. Advances in Research Ethics and Integrity* (2). Emerald: 79-107.

Association of Internet Researchers (2012). "[Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee \(Version 2.0\)](#)".

Batrinca, B. & Treleaven, P. C. (2015). "[Social media analytics: a survey of techniques, tools and platforms](#)". *AI & Society*, vol. 30(1): 89-116.

British Psychological Society (2017). "[Ethics Guidelines for Internet-mediated Research](#)".

British Sociological Society (2017). "[Ethics Guidelines and Collated Resources for Digital Research: Statement of Ethical Practice Annexe](#)". This resource contains individual case study documents:

1. [Research Online Forums](#)
2. [Using Twitter for Criminology Research](#)
3. [Researching Social Machines](#)
4. [New Methodological Approaches to Research Using Twitter](#)
5. [Mixed Methods Study Involving Young People](#)
6. [Open Data and Democratic Governance: Policies, Platforms and Practices](#)

Economic and Social Research Council (2019). "[Internet-mediated research](#)".

Golder, S., et al (2017), "[Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review](#)". *Journal of Medical Internet Research*, vol. 19(6): e195.

Information Commissioner's Office (2017). [Big data, artificial intelligence, machine learning and data protection](#), Version 2.2. (ICO: England).

Information Commissioner's Office (2018). ["GDPR Individual Rights: The Right to be Informed"](#).

Munson et al. (2013). ["Sociotechnical Challenges and Progress in Using Social Media for Health"](#). *Journal of Medical Internet Research* 15(10): e226.

Taylor, J., and Pagliari, C. (2018). ["Mining Social Media Data: How are Research Sponsors and Researchers Addressing the Ethical Challenges?"](#). *Research Ethics* 14(2): 1-39.

Townsend, L. & Wallace, C. (2017). [Social Media Research: A Guide to Ethics](#).

UK Research Integrity (2018). ["GDPR and Research: An Overview for Researchers"](#).

UK Research Integrity Office (2016). ["Good Practice in research: Internet-mediated research"](#).

Flowchart demonstrating considerations of ethics in research using social media data.

